

## Performance of a Diagnostic System (Iliad) as a Tool for Quality Assurance

LEE MIN LAU AND HOMER R. WARNER

*Department of Medical Informatics, University of Utah Medical Center,  
Salt Lake City, Utah 84132*

Received April 2, 1992

Quality assurance improves health care through detection of quality problems and feedback to the care giver. Current review procedures employed by the Peer Review Organizations (PROs), however, appear to underdetect quality problems, particularly those arising from diagnostic errors. We studied the use of an expert diagnostic system, Iliad, to detect quality problems arising from diagnostic errors. One hundred cases were selected from among those Medicare cases reviewed by the Utah PRO (UPRO) and which contained diagnoses recognized by Iliad. Iliad flagged 28 cases out of the 100 as containing diagnostic errors, and a gold standard physician review confirmed quality problems in 17 cases (60.7%). The UPRO review found 28 cases with quality problems, mostly treatment and documentation errors. The quality problems detected by Iliad appeared to be more serious than those detected by the UPRO review. Among the six cases with quality problems detected by both the UPRO and the Iliad review, there was none for which the same quality problem was detected by the two procedures. The two review procedures were therefore complementary. © 1992 Academic Press, Inc.

### INTRODUCTION

The goal of quality assurance is to improve health care delivery through the monitoring and analysis of patient management strategies (1-4). The detection of a quality problem leads to feedback to the care giver and thus improvement in future patient care. Ideally, the monitoring and analysis for quality problems is performed by expert physicians through peer review. However, the high cost of physician review combined with the great volume of patient cases screened by the PROs prohibits a physician review of all cases. Therefore, the PROs sample a subset of cases to go through a preliminary nurse screening. The nurse reviewers apply generic quality screening rules and flag cases containing potential quality problems, which are then referred for an expert physician review (5, 6). The UPRO reviews about 10,000 Medicare cases a year (approximately a quarter of all Utah Medicare claims) (7). From April 1989 to March 1990, 18% of the Medicare cases reviewed were flagged by the nurse reviewers as containing potential quality problems and referred to physician reviewers.

The physician review confirmed quality problems in only 5.5% of these 18%, or 1% of the original 10,000 cases. This is a surprisingly low figure, compared to the much higher base rate of quality problems indicated by a substantial body of medical literature, which ranges from 3 to 42% (8-12, 14-16). Other PROs have reported quality problem rates of between 0.3 to 4.6% (17), and the national PRO average was reported to be 1.57% (18).

The UPRO's lower problem detection rate, compared to those found by other researchers, may be due to the underdetection of quality problems arising from diagnostic errors by current PRO review procedures. The present PRO review begins with the identification of probable quality problems by the nurses. Only then will these quality problems be referred to physician reviewers, with specific questions from the nurses requiring answers from the physicians. Nurses do not receive specific training in diagnosis, and the generic quality screens used in the nurse review focus almost entirely on therapeutic and documentation errors. Because physicians key their treatment strategies to diagnosis, diagnostic errors can result in management and therapeutic errors, which may not be apparent unless the nurse realizes that the diagnosis is incorrect. In practice, the nurses usually assume that the diagnosis is correct. On average, UPRO nurses question the diagnosis on only about a dozen charts each year, out of the 10,000 charts reviewed (verbal communication with Dr. A. Lloyd Poulsen, medical director of UPRO). Previous research indicates that diagnostic errors are much more widespread than this low UPRO rate suggests (8-11). For instance, one study reported a missed or delayed diagnosis in 10% of a series of inpatient cases covering five DRGs in internal medicine (10).

Research has found that diagnostic errors, through their influence on subsequent patient management, adversely affect patient health outcome. Diagnostic errors may lead to delayed or inappropriate investigations or treatments (12-16, 19). For example, a study of 64 cases of acute myocardial infarction in which the diagnosis was initially missed showed that the patients experienced an 83% mortality rate (13), compared to an expected mortality rate of 25% (14).

A review of the literature thus suggests that current PRO review procedures may be underdetecting important quality problems arising from diagnostic errors. The detection of quality problems is the prerequisite to feedback and improvement in health care delivery. Increasing the problem detection rate by solely using physician peer review is unlikely to be practical, given the large number of cases to be reviewed. Expert systems (computerized diagnostic systems) may provide an effective alternative means to detect quality problems (20).

Currently, the UPRO is one of seven PROs preparing to test and implement a Uniform Clinical Data Set (UCDS). The UCDS is a rule-based expert system that contains the present generic quality screens from the Health Care Financing Administration (HCFA). The PRO nurse reviewers enter the clinical data as requested by the UCDS, which includes such items as vital signs, laboratory results, and procedures, but not many history or physical findings. The UCDS system will analyze the data and apply multiple HCFA quality screens. How-



ever, these screens are not specifically designed to detect diagnostic errors that may lead to quality problems (verbal communication with Dr. A. Lloyd Poulsen, medical director of UPRO). Stewart *et al.* (20) described a computerized quality assurance system to assist a full-time quality assurance officer. The system contained quality screening criteria for emergency room case review. The number of patient cases referred for investigation of questionable care rose from a preimplementation rate of five patient care errors per month to 35 per month (20).

We studied the potential use of a medical expert system, called Iliad, to detect diagnostic errors that lead to important quality problems in patient care. Iliad is designed to act as a diagnostic consultant in internal medicine (21). Given the same set of patient data, a difference in opinion between Iliad and the case physician regarding the diagnosis may indicate the presence of a diagnostic error, which may give rise to a quality problem in subsequent patient management. This paper described Iliad's performance in flagging potential problem cases, to determine whether Iliad may be used as a supplementary screening tool for the current PRO review.

#### METHODS AND PROCEDURES

Iliad is an expert system designed for diagnosis in internal medicine (21). Currently, Iliad recognizes over 5000 medical findings and 1100 diagnostic conditions in nine subspecialties of internal medicine. Iliad's knowledge base is organized into frames, which employ probabilistic as well as rule-based logic. The knowledge base is continually being expanded and evaluated in an ongoing knowledge engineering effort.

The UPRO reviewed approximately 10,000 Medicare inpatient cases during 1989 and a paper copy of those charts flagged by the nurses as containing a potential quality problem was kept in the UPRO. A list of the 1100 diagnoses contained in Iliad was given to the UPRO, who retrieved 242 inpatient charts containing one of these diagnoses from among its in-house paper charts. Each of these charts had been reviewed by a nurse, found to have contained at least one potential quality problem, and referred to a physician. These cases were selected because we hypothesized that Iliad would detect a different type of error from that focused on by the nurse review. One hundred charts were randomly selected from among these 242 charts for use in our experiment.

Iliad's "consultation" mode was used in this experiment. In this mode, the case findings were entered into Iliad by either typing a keyword to bring up the corresponding data item for confirmation (for example, typing "fever" to bring up "present history of fever") or selecting from a list of data items contained in Iliad. Iliad would then provide a list of differential diagnoses which explained the findings. For the experiment, a general practitioner reviewed the patient record and entered the case findings obtained from history, physical examina-



tion, and investigative laboratory tests prior to treatment into Iliad. Iliad then generated a list of the top 20 differential diagnoses, ranked according to the diagnostic certainty (probability) Iliad assigned to each diagnosis.

The attending physician's discharge diagnoses from the discharge summary were compared to Iliad's diagnoses for the case. Postoperative complications as well as chronic conditions which were not the cause for admission and did not manifest during the hospitalization were not included. This was because Iliad would not diagnose postoperative complications and chronic conditions with no acute manifestation, since Iliad only used the data obtained during the admission diagnostic workup. An example of a chronic condition would be hypertension. Iliad would not diagnose hypertension if the condition was known and treated before admission, and the patient remained normotensive during the course of the present illness.

A potential diagnostic error was identified when there was a discrepancy between the attending physician's list of discharge diagnoses and the corresponding list provided by Iliad for the case. A discrepancy was defined as one of the following conditions:

- A diagnosis on the physician's list did not reach above 20% probability on Iliad's list ("unlikely diagnosis").
- A diagnosis exceeding 80% probability on Iliad's list did not appear on the physician's list ("missed diagnosis").

An identified discrepancy between the physician's diagnostic list and Iliad's would flag the case as requiring physician review. The data entry for all 100 cases were completed before the diagnostic lists were tabulated and compared for discrepancies.

Each of the charts flagged by a discrepancy between the physician's and Iliad's diagnostic list was submitted to an expert from the appropriate subspecialty of internal medicine for a "gold standard" review. Iliad's diagnoses were not revealed to the expert. Instead, specific questions arising from the discrepancy were asked. For instance, if the physician diagnosed unstable angina but not acute myocardial infarction, and Iliad diagnosed acute myocardial infarction with a probability of 87%, the expert reviewer would be asked if the patient could have suffered from an acute myocardial infarction. After the expert has reviewed the case, if he replied in the affirmative, he would be asked if there was then a quality problem in the management of the case, and if so, to rate the severity of the quality problem by the following nationwide standard PRO quality classification. Using the standard PRO weighting criteria, a severity score was also assigned to each case according to the level of the quality problem (6):

- Care appropriate—without quality problems.
- Level I quality problem (severity score 1)—without potential for significant adverse effects on the patient.
- Level II quality problem (severity score 5)—with potential for significant adverse effects on the patient.

- Level III quality problem (severity score 25)—with significant adverse effects on the patient.

## RESULTS

The ages of the 100 patients ranged from 31 to 102 years. The average age was 76, with a standard deviation of 11 years. Forty-one of the 100 patients were female. The earliest admission date was March 30, 1989, and the latest discharge date was May 28, 1990. The length of stay ranged from 2 to 27 days, with an average of 8 days and a standard deviation of 5 days. There were five deaths among the 100 cases. The admissions were to 22 hospitals in Utah, and there were 31 different principal discharge diagnoses among the 100 cases. The most common principal diagnosis appeared to be unstable angina (28 cases), followed by pneumonia (18 cases) and benign prostatic hypertrophy (6 cases).

*UPRO review.* The investigators were blinded to the results of the UPRO review until the completion of the gold standard review for the cases flagged by Iliad. All 100 cases were flagged by the UPRO nurse review as containing quality problems and referred for physician review. The results of the UPRO physician review are summarized in Table 1.

The 28 confirmed quality problems were:

Level I problem: 21 cases

Inadequate discharge planning: 3

Medical stability not ensured at discharge: 9

Medication error: 1

Failure to perform a test: 1

Inadequate documentation: 5

Failure to perform a physical examination: 1

Nosocomial infection: 1

Level II problem: 7 cases

Medical stability not ensured at discharge: 4

Failure to perform a test: 2

Medication error: 1

The nurses required 5 to 40 min per review, with an average of 20 min. The physician review averaged 36.7 min, ranging from 15 to 153 min.

*Iliad review.* Twenty-eight out of the 100 cases were flagged by Iliad for the "gold standard" review, using the previously described criteria for discrepancy. Each case was reviewed by an expert physician in the appropriate subspecialty of internal medicine. The results of the physician review are summarized in Table 2.

The 17 confirmed quality problems were:

Level I problem: 10 cases

Missed acute myocardial infarction (AMI): 1

Missed osteoarthritis: 1

Missed severe secondary hyperparathyroidism: 1



TABLE 1  
RESULTS OF THE UPRO REVIEW

Appropriate care	72
Quality problems	<b>28</b>
Level I (score 1)	21
Level II (score 5)	7
Level III (score 25)	0
Average severity score of confirmed problems	2.0
Cases flagged by nurses	<b>100</b>
% of flagged cases confirmed to have problems	28.0

Unlikely diagnosis of chronic lymphocytic leukemia: 1

Unlikely diagnosis of acute gastritis: 1

Unlikely diagnosis of peritonitis: 1

Unlikely diagnosis of AMI: 2

Unlikely diagnosis of diverticulitis: 2

Level II problem: 5 cases

  Missed AMI: 1

  Missed diabetes mellitus: 2

  Missed diabetic nephropathy: 1

  Missed severe secondary hyperparathyroidism: 1

Level III problem: 2 cases

  Missed pulmonary embolus: 1

  Missed lung cancer: 1

Interestingly, a missed diagnosis by the attending physician appeared to result in more serious quality problems than if an unlikely diagnosis was made (see Table 3).

Fifteen to 40 min was needed to enter the data from one patient record into Iliad, with an average of 29.2 min. The "gold standard" physician review required 2 to 20 min per case, averaging 8.8 min.

*Comparing UPRO and Iliad Reviews.* The results of the UPRO (nurse-physician) review and the Iliad (Iliad-"gold standard" physician) review were com-

TABLE 2  
RESULTS OF THE ILIAD REVIEW

Appropriate care	11
Quality problems	<b>17</b>
Level I (score 1)	10
Level II (score 5)	5
Level III (score 25)	2
Average severity score of confirmed problems	5.3
Cases flagged by Iliad	<b>28</b>
% of flagged cases confirmed to have problems	60.7

TABLE 3  
QUALITY PROBLEMS DETECTED BY THE ILIAD REVIEW:  
MISSED DIAGNOSES VERSUS UNLIKELY DIAGNOSES

Quality problem	Level I (score 1)	Level II (score 5)	Level III (score 25)	Total
Missed diagnosis	3 (30%)	5 (50%)	2 (20%)	10 (100%)
Unlikely diagnosis	7 (100%)	0 (0%)	0 (0%)	7 (100%)

pared to determine whether the same quality problems were detected by the two independent processes. The results of this comparison are summarized in Table 4.

Among the six cases with quality problems detected by both the UPRO and the Iliad review, there was none for which the same quality problem was detected by the two procedures. That is, the UPRO review and Iliad detected quality problems of different nature. For instance, in one of these six cases, UPRO review found that the consent for a procedure was not in the case record. Iliad review, however, found that a diagnosis of AMI was missed.

#### DISCUSSIONS AND CONCLUSIONS

The UPRO review of the 100 inpatient Medicare cases found 28 cases (28.0%) that contained quality problems. This is much higher than the 1% problem rate reported in the 1990 UPRO annual report (6), which was, however, derived from all the cases reviewed by UPRO across the specialties of medicine, including surgical and psychiatric disciplines. Our sample of 100 cases was restricted to those with a principle diagnosis Iliad was able to recognize, mainly problems in the field of internal medicine. We therefore only evaluated Iliad's performance on cases that fell within its domain of expertise.

Twenty-eight of the 100 cases were flagged by Iliad as requiring physician

TABLE 4  
COMPARING UPRO AND ILIAD REVIEWS

UPRO review	Iliad review		Total
	Quality problem detected	Quality problem not detected	
Quality problem detected	6	22	28
Quality problem not detected	11	61	72
Total	17	83	100



TABLE 5  
COMBINING UPRO AND ILIAD REVIEWS

Review process	UPRO	Iliad	Combined
% of flagged cases confirmed to have quality problems	28.0	60.7	39.0
Average severity score of confirmed quality problems	2.0	5.3	3.4
Nurse review time (min)	20.0	29.2	49.2
Physician review time (min)	36.7	8.8	45.5

review. The "nurse review time" in the following table, when applied to the Iliad review, refers to the time required to enter data from a case record into Iliad. In our experiment, the data entry was performed by a general practitioner, but the procedure could be performed by a nurse. Seventeen of the 28 flagged cases were confirmed by the "gold standard" physician review to have contained quality problems (60.7%). As expected, the quality problems detected by the physician reviewers in the cases flagged by Iliad were different in nature from those detected by the UPRO review. Since there was no overlap between the types of quality problems detected by the UPRO and the Iliad reviews (see Table 4), the total number of quality problems detected by using both procedures would be  $(6 + 11 + 22)$ , or 39, of the 100 cases reviewed. In addition to increasing the number of quality problems found, using Iliad in addition to the PRO review would result in more serious quality errors being found. The average score of the confirmed quality problems detected by Iliad review was 5.3, higher than the average score of the quality problems detected by the UPRO review, which was 2.0 (see Tables 1 and 2). More importantly, Iliad review resulted in the detection of two level III problems not detected by the PRO review (see Tables 1 and 2). These level III quality problems resulted in significant adverse effect on the patient (see Table 5).

Assuming that the cost of nurse review is \$15.00 an hour, and that of physician review \$60.00 an hour, the cost per quality problem confirmed can be calculated for the UPRO and Iliad reviews (see Table 6).

Although using Iliad in addition to standard PRO review resulted in an increase in net cost, the cost per quality problem confirmed actually decreased.

Among the 11 cases found to have received appropriate care by the Iliad physician review, Iliad's diagnoses were judged to be incorrect in nine cases. The nine false positives flagged by Iliad were analyzed in our knowledge engineering sessions after the experiment. In two of the cases, there was a mistake in the data entry which resulted in Iliad's generation of a different diagnostic list from that of the attending physician. For the remaining seven cases, Iliad was at fault, and the relevant disease frames in Iliad were corrected. Therefore our experiment also provided a useful way of evaluating and improving Iliad's diagnostic accuracy. In our experiment, the version of Iliad used for the first



TABLE 6  
COSTS OF UPRO AND ILIAD REVIEWS

Review process	UPRO	Iliad	Combined
Cost of nurse review	\$500.00 (100 cases)	\$730.00 (100 cases)	\$1230.00 (100 cases)
Cost of physician review	\$3670.00 (100 cases)	\$246.00 (28 cases)	\$3916.00 (100 cases)
Total cost	\$4170.00	\$976.00	\$5146.00
Number of problem cases confirmed	28	17	39
Cost per confirmed problem case	\$148.93	\$57.41	\$131.95

case was used for all 100 cases. However, the improved version of Iliad could be used for future review.

#### ACKNOWLEDGMENTS

The authors thank the Utah Peer Review Organization for its help and cooperation, especially Dr. Lloyd Poulsen, medical director; and Mr. Robert Vaughn, senior health data analyst. Special thanks go to Dr. David Bjorkman, Dr. Bruce Bray, Dr. Peter Haug, Dr. Michael Lincoln, Dr. Attilio Renzetti, and Dr. Frank Tyler who performed case review in our experiment. This study is supported in part by Grant 5R01-LM-046043 from the National Library of Medicine, and by Dissertation Grant 1-R03-HS06947-01 from the Agency for Health Care Policy and Research.

#### REFERENCES

1. VANAGUNAS, A., EGELSTON, E. M., HOPKINS, J., *et al.* Principles of quality assurance. *QRB* 5, 4 (1979).
2. VANAGUNAS, A. A. Quality assessments: Alternative approaches. *QRB* 5, 8 (1979).
3. JESSEE, W. F. Quality assurance systems: Why aren't there any? *QRB* 10, 408 (1984).
4. DONABEDIAN, A. Quality assessment and assurance: Unity of purpose, diversity of means. *Inquiry* 25, 173 (1988).
5. MOORE, S., RYAN, C. P., AND MURPHY, R. A. PRO review part III: Quality review, intervention, and sanction. *J. Med. Assoc. Ga* 78, 683 (1989).
6. MURPHY, R. A., WILLIAMS, T. W., AND RYAN, C. P. Peer review in the 1990s: A look at the Georgia medical care foundation. *J. Med. Assoc. Ga* 78, 549 (1989).
7. UTAH PEER REVIEW ORGANIZATION. Annual report to physicians and health care facilities. Review results April, 1989–March, 1990. (1990).
8. MERCER, J., AND TALBOT, I. C. Clinical diagnosis: A post-mortem assessment of accuracy in the 1980s. *Postgrad. Med. J.* 61, 713 (1985).
9. ANDERSON, E. A., ROLLA, B. H., AND KEY, C. R. The sensitivity and specificity of clinical diagnostics during five decades. *JAMA* 261, 1610 (1989).
10. WEINBERG, N. S. The relation of medical problem solving and therapeutic errors to disease categories. *QRB* 15, 266 (1989).
11. WILLIAMSON, J. W., FEHLAUER, C. S., *et al.* Assessing quality of ambulatory care. *Quality Assurance Utilization Rev.* 6, 8 (1991).
12. KASSIRER, J. P., AND KOPELMAN, R. I. Cognitive errors in diagnosis instantiation, classification, and consequences. *Am. J. Med.* 86, 433 (1989).

13. PELBERG, A. L. Missed myocardial infarction in the emergency room. *Qual. Assurance Utilization Rev.* **4**, 39 (1989).
14. WILSON, J. D., BRAUNWALD, E., et al. "Harrison's Principles of Internal Medicine," 12th ed. McGraw-Hill, New York, 1991.
15. ADAMS, I. D., CHAN, M., COOKE, W. M., et al. Computer-aided diagnosis of acute abdominal pain: A multicentre study. *Br. Med. J.* **293**, 800 (1986).
16. CHASSIN, R., KOSECOFF, J., PARK, R. E., et al. Does inappropriate use explain geographic variations in the use of health care services? A study of three procedures. *JAMA* **258**, 2533 (1987).
17. Annual reports: Nevada Peer Review, 1990; Puerto Rico Foundation for Medical Care, 1990; California Medical Review, 1990; New Mexico Medical Review Association, 1991; Island Peer Review Organization (New York), 1990; Medical Society of Virginia, 1991; Arkansas Foundation for Medical Care, 1990; Peer Review Systems (Ohio), 1990; Colorado Foundation for Medical Care, 1990; Iowa Foundation for Medical Care, 1990; Michigan Peer Review Organization, 1990.
18. PRO/W Annual report for Alaska, Idaho, and Washington, 1991.
19. BROOK, R. H., KAMBERG, C. J., et al. Appropriateness of Acute Medical Care for the Elderly—An Analysis of the Literature. R3717-AARP/HF/RWJ/RC. Rand Corp. Publishers, Santa Monica, 1989.
20. STEWART, R. D., BURGMAN, J., CANNON, G. M., et al. A computer-assisted quality assurance system for an emergency service. *Ann. Emergency Med.* **14**, 25 (1985).
21. WARNER, H. R., HAUG, P., BOUHADDOU, O., et al. Iliad As An Expert Consultant to Teach Differential Diagnosis. *SCAMC* **12**, 371 (1988).